

Graphen und biologische Netze Praktikum

Robin Goldmann, Johann Wurz, Christina Kuhn

Bewertung der EC-Klassifikation anhand des Vergleich der Substrate - ein Versuch

EC Database → PubChem → Json to Graph → Alignments → Konsensus-Sequenz

Substrate loader

- ECtoJSONLoader.py
- Input:
 - EC Subtree
 - Anzahl gewünschter Substrate pro Subtree
 - Maximale Anzahl an Knoten pro Substrat (ohne Wasserstoff -Atome)
- Output
 - Substrate im JSON Format

Substrate loader → Probleme

- unterschiedliche Substrate genannt → häufigstes genommen
- mehrere Substrate, die vom Enzym katalysiert werden
 - Filtern nach TTP, UTP, ITP, GTP, CTP, ADP, CoA, acetate, H₂O, ATP, CO₂, NAD⁺, NADPH, H₂, NAD(P)H, H⁺, O₂, NADH, NADHX, HCN
 - wenn nach dem Filtern immernoch mehrere Substrate → exclude subtree
- manche Ordner bleiben leer, da zu diesem Subtree keine Substrate gefunden werden konnten → Gründe:
 - Graphen zu groß
 - nur mehrere Substrate enthalten
 - Subtree eigentlich leer, da Substrat in andere Klasse geschoben wurde

Substrate loader → Ergebnisse

Anzahl der Subtrees

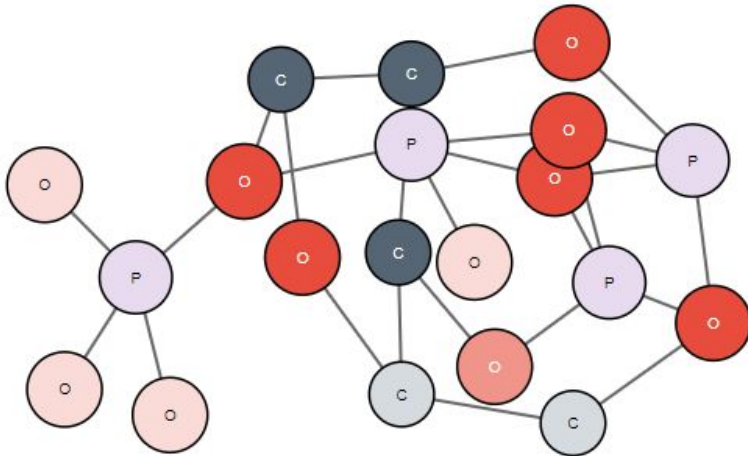
- EC 1.* → 302/4584 (6,6%) Oxidoreduktasen
- EC 2.* → 177/2023 (8.5%) Transferasen
- EC 3.* → 316/1621 (19.5%) Hydrolasen
- EC 4.* → 264/2419 (10.9%) Lyasen
- EC 5.* → 128/320 (40%) Isomerasen
- EC 6.* → 46/221 (20.8%) Ligasen
- EC 7.* → 9/70 (12.9%) Translokasen

Alignment Workflow

- mit Nextflow
-
1. Json to Graph (mit GraphDating)
 2. Exclude H-Atome (mit GraphDating -nh)
 3. Align Graphs mit Visualization (mit MultiVitamin -a subVF2 -ri <INPUT>)
 - a. nur Exakte Matches erlaubt: C-C, O-O, N-N

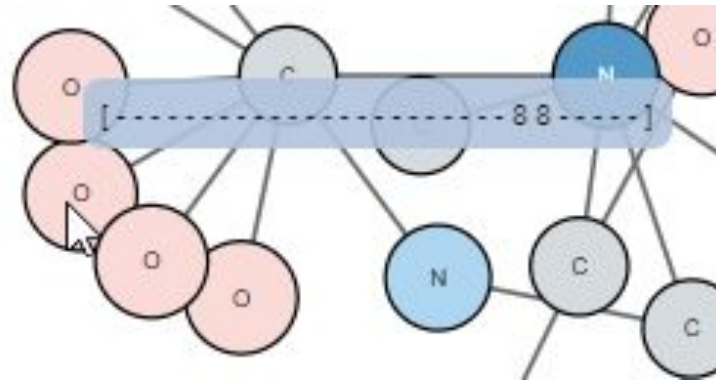
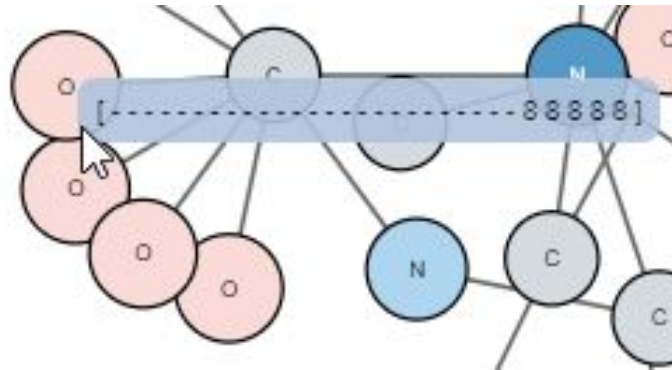
Alignment Workflow Probleme

- MultiVitamin ist abgebrochen
 - Fatal error while scoring, tried to score a matching of zero-length nodes
- MultiVitamin hat zu lange gedauert
 - Bsp.: 10 min für 5.4.2 → Idee: mit Guide Tree (UPGMA density) ca. 1 Sekunde

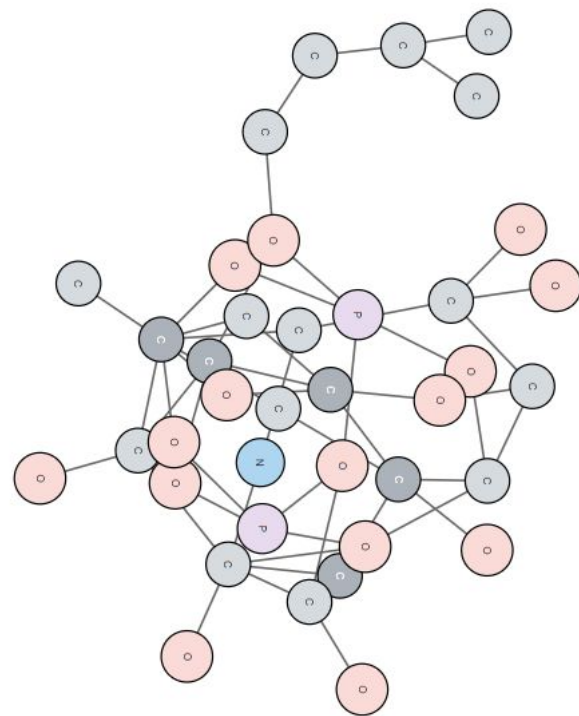
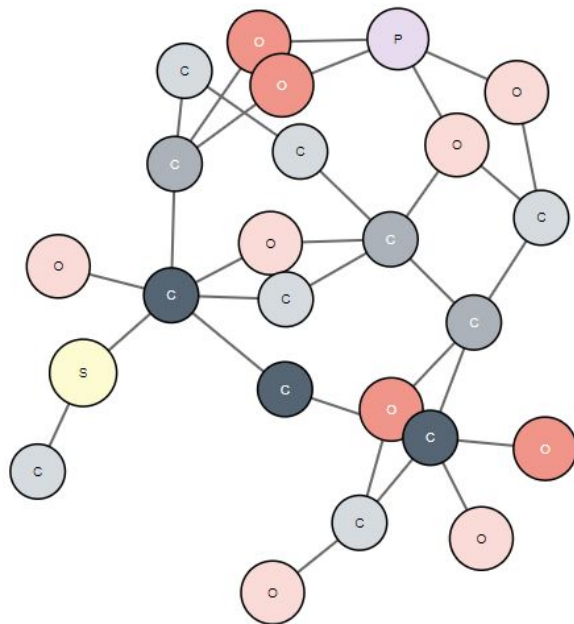
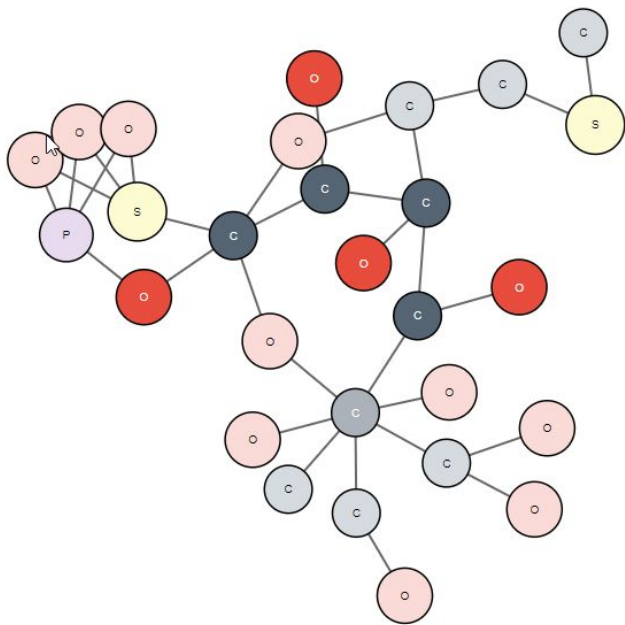


Alignment Analyse Probleme

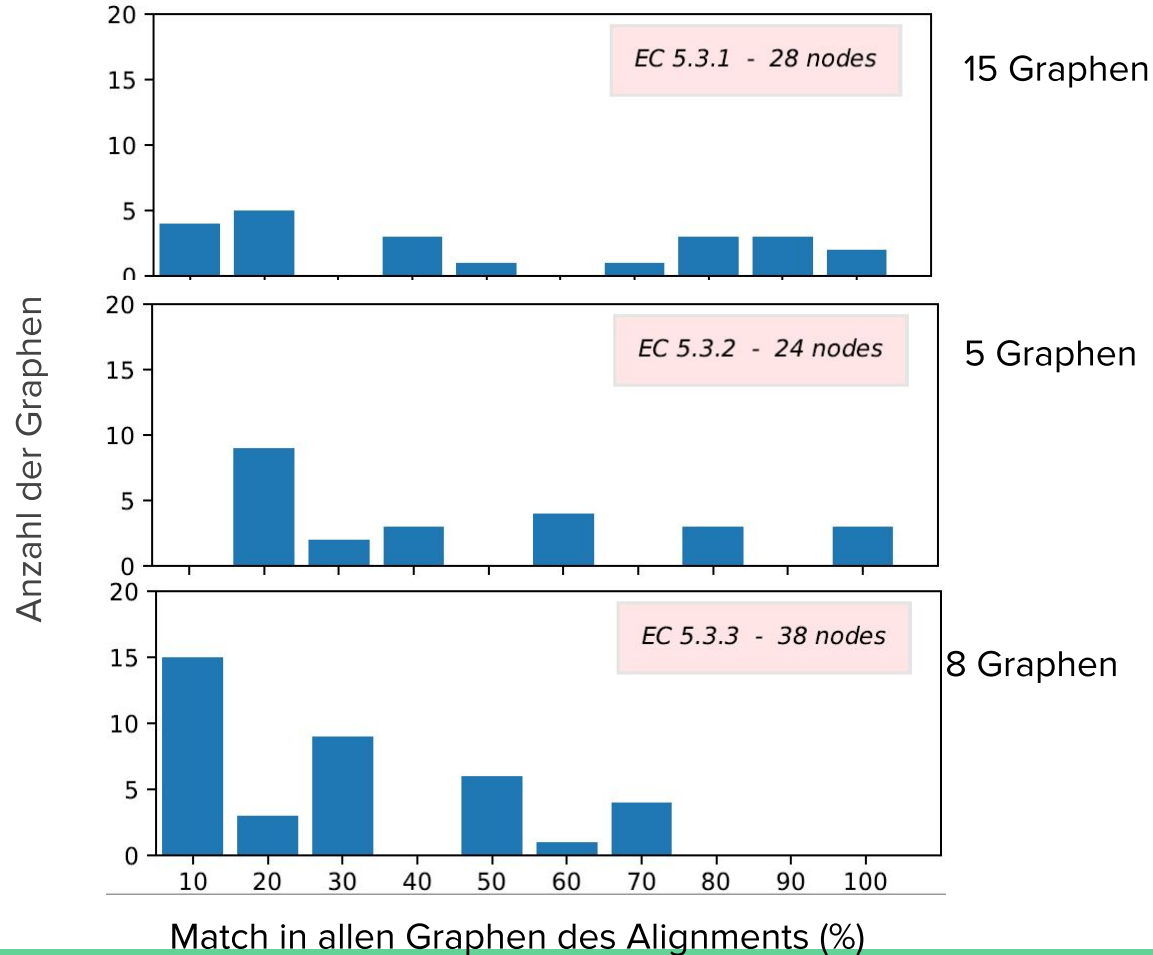
- An den Ränder häufig Knoten die gematched werden sollten, aber es nicht werden



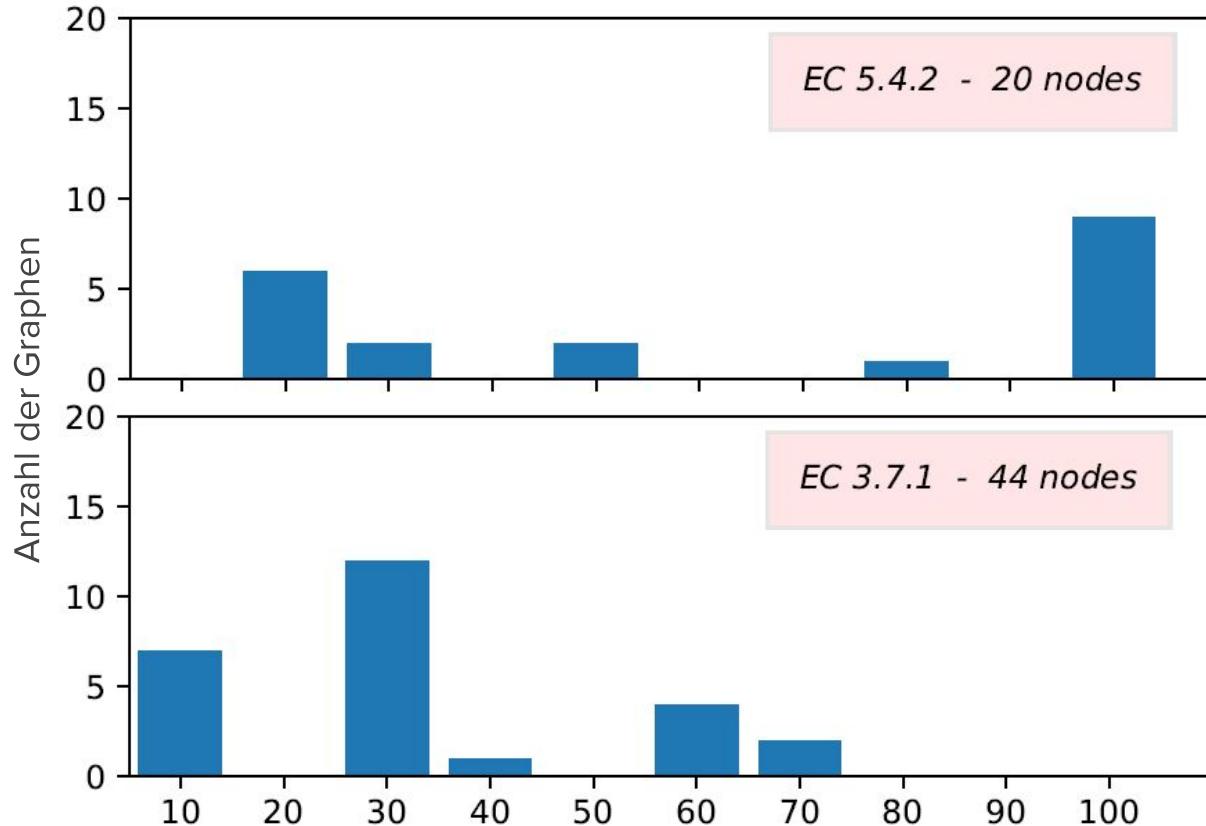
Alignment Analyse 5.3.1 vs. 5.3.2 vs. 5.3.3



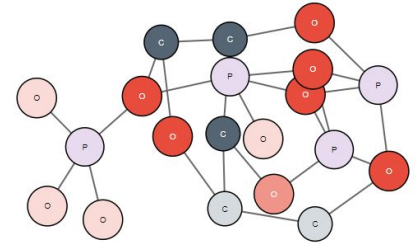
Alignment Analyse 5.3.1 vs. 5.3.2 vs. 5.3.3



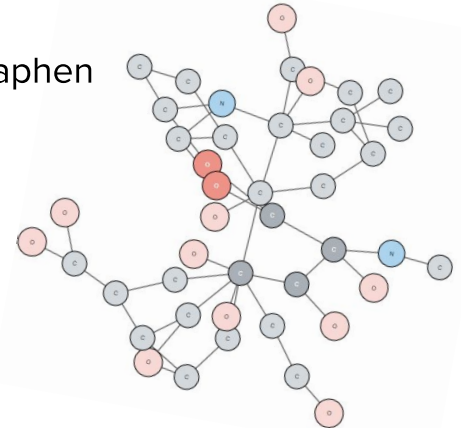
Alignment Analyse Best vs. Worst



5 Graphen



13 Graphen



Zusammenfassung

